



Bayesian bivariate survival estimation

J.K. Ghosh, N.L. Hjort, C. Messan, R.V. Ramamoorthi*

Department of Statistics and Probability, Michigan State University, Wells Hall, East Lansing, MI 48824-1027, USA

Available online 6 September 2005

Abstract

There is no easy extension of the Kaplan–Meier and Nelson–Aalen estimators to the bivariate case, and estimating bivariate survival distributions nonparametrically is associated with various nontrivial problems. The Dabrowska estimator will, for example, associate negative mass to some subsets. Bayesian methods hold some promise as they will avoid the negative mass problem, but are also prone to difficulties. We simplify and extend an example by Pruitt to show that the posterior distribution from a Dirichlet process prior is inconsistent. We construct a different nonparametric prior via Beta processes and provide an updating scheme that utilizes only the most relevant parts of the likelihood, and show that this leads to a consistent estimator.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Bayesian nonparametrics; Beta processes; Bivariate survival; Inconsistency

1. Introduction

Bivariate survival or waiting times to occurrence of some event has several interesting applications. For example, a husband and his wife may both be exposed to a common risk and one might want to know if males and females have the same survival distribution. Another situation might be the study of effects of active and passive smoking when only one of the spouses smokes.

In the univariate case, the most popular nonparametric estimator is the Kaplan–Meier one. The bivariate case remains surprisingly difficult in spite of a lot of work. Dabrowska (1988) constructs a bivariate analog of the Kaplan–Meier estimate which is consistent but is not a proper survival distribution in that it assigns negative mass to some events. The same is true of an earlier estimate of Langberg and Shaked (1982). Dabrowska (1988) mentions another estimate due to Bickel which is consistent and avoids negative masses but does not seem to utilize all of the data. Pruitt (1991) shows mathematically that the negative mass assigned by Dabrowska's estimator can be

* Corresponding author.

E-mail address: ramamoorthi@stt.msu.edu (R.V. Ramamoorthi).

substantial even though the estimate assigns positive mass to all upper orthants. A fairly complete survey of some of these approaches can be found in Andersen et al. (1993, Chapter X).

Pruitt (1988) also shows that the Bayes estimate of the bivariate survival function with a Dirichlet process prior can be dramatically inconsistent. In Pruitt (1993), he suggests an independence assumption weaker than Dabrowska's and explains how in this context one can construct a new bivariate Kaplan–Meier estimate which is a proper survival function.

The univariate Kaplan–Meier estimate can be obtained as a limit of Bayes estimates with respect to Dirichlet process priors, more generally Beta processes (see Hjort, 1990) and even more generally through neutral to the right priors (Hjort, 1990). In this paper, we develop a natural generalization of Beta processes for the bivariate case and derive an estimator using an essentially Bayesian approach. The qualification 'essentially' is used since we do not use the full likelihood.

The organization of this paper is as follows. In Section 2, we discuss Pruitt's example and give a simple proof of inconsistency. In Section 3, we consider discrete distributions for the survival time and through a convenient parametrization relate it to the distribution of the censored data. The discussion in this section provides some insight into the difficulties associated with the bivariate case. Then in Section 4, we look at the time-discrete case and develop the Bayesian analysis. The last section considers the general case. Though we restrict ourselves to Beta processes, much of our arguments and results would go through for neutral to the right priors, but we do not pursue this here.

2. Pruitt's example

We follow Pruitt (1988) but provide a rather easy proof of inconsistency. Let (T_1, T_2) be two survival times, (C_1, C_2) the censoring variables, and let $Z_j = T_j \wedge C_j$ and $\Delta_j = I\{T_j \leq C_j\}$ for $j = 1, 2$ be the observed random variables.

Let (C_1, C_2) be discrete with mass $\frac{1}{3}$ at each of $(1, 3)$, $(3, 1)$, $(4, 4)$. The (T_1, T_2) pair is assumed to take values in $\mathcal{T} = [1, 3] \times [1, 3]$. The prior is a Dirichlet process with parameter $M\alpha$, where α is the uniform distribution on $[1, 3] \times [1, 3]$. We will show that if the true distribution P_0 is uniform on $A = [1, 2] \times [1, 2] \cup [2, 3] \times [2, 3]$ then the posterior is inconsistent. Much more delicate calculations are needed to show that the estimate converges weakly to $(\frac{1}{3})P_0 + (\frac{2}{3})\alpha$. Let $(\tilde{Z}, \tilde{\Delta})$ denote the observations $(Z_{1,i}, \Delta_{1,i}), (Z_{2,i}, \Delta_{2,i})$ for $i = 1, \dots, n$, with

$$\tilde{Z}_i = (Z_{1,i}, Z_{2,i}) \quad \text{and} \quad \tilde{\Delta}_i = (\Delta_{1,i}, \Delta_{2,i}).$$

Consider $B = [1, 2] \times [2, 3]$. Then $P_0(B) = 0$. Let $E\{P(B) | (\tilde{Z}, \tilde{\Delta})\}$ be the Bayes estimator. We shall show that this does not go to 0 in probability as $n \rightarrow \infty$.

Recall that if E is a subset of $\mathcal{T} = [1, 3]$ then with a Dirichlet(α) prior for the distributions on $\mathcal{T}$, the posterior given E is the mixture

$$\int D_{\alpha + \delta(\tilde{t})} \alpha(d\tilde{t} | E),$$

where $\delta(\tilde{t}) = \delta_{\tilde{t}}$ denotes unit mass at point $\tilde{t}$, i.e. $\delta(\tilde{t})(B) = I\{\tilde{t} \in B\}$, and $\alpha(\cdot | E)$ is the conditional distribution of $\tilde{t}$ given E . Extending this to n observations, and thinking of the censored observations as specifying subsets of $\mathcal{T}$, for instance $(\Delta_1, \Delta_2) = (0, 1)$ as corresponding to the

horizontal line at Z_1 , the posterior is a mixture of the form

$$\int D_{\alpha+\sum_{i=1}^n \delta(\tilde{t}_i)} \alpha(d\tilde{t}_i | (\tilde{Z}_1, \tilde{A}_1), \dots, (\tilde{Z}_n, \tilde{A}_n)).$$

Consequently, the Bayes estimate of P is

$$\frac{M}{M+n} \alpha + \frac{n}{M+n} \frac{1}{n} \sum_{i=1}^n E\{\delta(\tilde{t}_i) | (\tilde{Z}, \tilde{A})\}.$$

Since P_0 is continuous, we may assume that $\tilde{t}_1, \dots, \tilde{t}_n$ are all distinct. Given this assumption, $\tilde{t}_1, \dots, \tilde{t}_n$ are independent with common distribution α . Hence, the Bayes estimate of $P(B)$ is

$$\frac{M}{M+n} \alpha(B) + \frac{n}{M+n} \frac{1}{n} \sum_{i=1}^n E[\delta(\tilde{t}_i)(B) | ((Z_{1,i}, A_{1,i}), (Z_{2,i}, A_{2,i}))].$$

The expectation in the above expression is calculated taking $\tilde{t}_1, \dots, \tilde{t}_n$ to be independently drawn from α . As $n \rightarrow \infty$, the first term in the last expression goes to 0 so we shall concentrate on the second term.

Note that under the model it is not possible to have both A_1 and A_2 equal to 0. Further, if $A_{1,i} = 1, A_{2,i} = 1$ for some i then the conditional distribution of $\tilde{t}_i$ given $(Z_{1,i}, A_{1,i}), (Z_{2,i}, A_{2,i})$ is degenerate at $(Z_{1,i}, Z_{2,i})$. So observations with $A_{1,i} = 1, A_{2,i} = 1$ contribute to the Bayes estimate of $P(B)$ only if the corresponding $(Z_{1,i}, Z_{2,i})$ is in B . Since $P_0(B) = 0$, under P_0 there will be no such observations and so will not contribute to the Bayes estimate of $P(B)$. Thus, the only case that we need to consider is when exactly one of $A_{1,i}$ and $A_{2,i}$ is 1 and the other is 0. Thus,

$$\begin{aligned} & \sum_{i=1}^n E\{\delta_{\tilde{t}_i}(B) | (Z_{1,i}, A_{1,i}), (Z_{2,i}, A_{2,i})\} \\ &= \sum_{i: A_{1,i}=0, A_{2,i}=1} E\{\delta_{\tilde{t}_i}(B) | (Z_{1,i}, A_{1,i}), (Z_{2,i}, A_{2,i})\} \\ &+ \sum_{i: A_{1,i}=1, A_{2,i}=0} E\{\delta_{\tilde{t}_i}(B) | (Z_{1,i}, A_{1,i}), (Z_{2,i}, A_{2,i})\}. \end{aligned}$$

Next, consider an i for which $Z_{1,i} = z_{1,i}, Z_{2,i} = z_{2,i}$ and $A_{1,i} = 0, A_{2,i} = 1$. In this case, we know that $T_1 \in [1, 3]$ and $T_2 = z_{2,i}$. Consequently, under α , the conditional distribution of T_1 given $\{Z_{1,i} = z_{1,i}, A_{1,i} = 0, Z_{2,i} = z_{2,i}, A_{2,i} = 1\}$ is uniform on $[1, 3]$. From these observations it follows that if $Z_{2,i} \in [2, 3]$,

$$E_{\tilde{t}_i}\{\delta_{\theta_i}(B) | Z_{1,i}, A_{1,i} = 0, Z_{2,i}, A_{2,i} = 1\} = E\{I_B(T_1, T_2) | T_{1,i} \geq 0, T_{2,i} = Z_{2,i}\}$$

and equal to zero otherwise. Now, if $Z_{2,i} \in [2, 3]$ then

$$E\{I_B(T_1, T_2) | T_{1,i} \geq 0, T_{2,i} = Z_{2,i}\} = \frac{1}{2} \int_1^2 dt_1 = \frac{1}{2}.$$

One handles the case of $\Delta_{1,i} = 1$, $\Delta_{2,i} = 0$ similarly. Hence,

$$\begin{aligned} & \frac{n}{M+n} \frac{1}{n} \sum_{i=1}^n E\{\delta_{\tilde{T}_i} \mid (Z_{1,i}, \Delta_{1,i}), (Z_{2,i}, \Delta_{2,i})\} \\ &= \frac{n}{M+n} \left\{ \frac{1}{2n} \sum_{i:\Delta_{1,i}=0, \Delta_{2,i}=1} I_{[2,3]}(Z_{2,i}) + \frac{1}{2} \sum_{i:\Delta_{1,i}=1, \Delta_{2,i}=0} I_{[1,2]}(Z_{1,i}) \right\} \end{aligned}$$

and as $n \rightarrow \infty$ this goes to

$$\left(\frac{1}{2}\right)[P\{Z_2 \in [2, 3], \Delta_1 = 0, \Delta_2 = 1\} + P\{Z_1 \in [1, 2], \Delta_1 = 1, \Delta_2 = 0\}].$$

Using the fact that

$$P\{Z_2 \in [2, 3], \Delta_1 = 0, \Delta_2 = 1\} = P\{Z_1 \in [1, 2], \Delta_1 = 1, \Delta_2 = 0\} = \left(\frac{1}{2}\right)\left(\frac{1}{3}\right),$$

the limit is seen to be $\frac{1}{6}$. This shows that the Bayes estimator is inconsistent and hence so is the posterior.

It might appear that the difficulty stems from P_0 and the prior guess not having the same support. However, a little reflection will show that the inconsistency phenomenon will occur even if P_0 and the prior guess have the same support.

3. Parametrization of distributions

Before we proceed to the details it is convenient to have an overview of the developments in this section.

Consider the usual one-dimensional case, where T is a life-time with unknown distribution F and C , independent of T , is a censoring variable with a fixed known distribution G . The observation is (Z, Δ) , where $Z = T \wedge C$ and $\Delta = I\{T \leq Z\}$. Let $\mathcal{M}(T)$, $\mathcal{M}(Z, \Delta)$ denote, respectively, the set of all distributions of T and (Z, Δ) . Under minimal assumptions the map $\phi: \mathcal{M}(T) \mapsto \mathcal{M}(Z, \Delta)$, which takes any F to $F^{(Z, \Delta)}$, is one-one and onto. In particular, Peterson (1977) has shown that the inverse of the empirical distribution leads to the Kaplan–Meier estimate.

The situation is drastically different in the two- and higher-dimensional case. Here, if $\mathcal{M}(\tilde{T})$, $\mathcal{M}(\tilde{Z}, \tilde{\Delta})$ denote, respectively, the set of all distributions of $\tilde{T}$, $(\tilde{Z}, \tilde{\Delta})$, the corresponding map $\phi: \mathcal{M}(\tilde{T}) \mapsto \mathcal{M}(\tilde{Z}, \tilde{\Delta})$, which takes any F to $F^{(\tilde{Z}, \tilde{\Delta})}$, is one-one but far from onto. In fact, many empirical distribution functions fall outside the range. What we do in this section is, in an abstract sense, to define an ‘inverse’ to all of $\mathcal{M}(\tilde{Z}, \tilde{\Delta})$, and we do this through a reparametrization of the underlying distributions.

We first reparametrize the distribution of T , then carry out a similar operation for the distribution of $(\tilde{Z}, \tilde{\Delta})$ and show that using only part of the distribution of $(\tilde{Z}, \tilde{\Delta})$ we can recover the distribution of T . This makes it transparent that the function ϕ is not onto. We then use the reparametrization to write the likelihood function. We note that it contains a component that is complex and propose that it be ignored. It will be seen in the next section that working with this incomplete likelihood enables one to develop a tractable posterior.

Let $\tilde{T} = (T_1, T_2)$ have the distribution P , a discrete probability measure on the positive quadrant, for concreteness and without essential loss of generality taken to have support in $\{1, 2, 3, \dots\} \times \{1, 2, 3, \dots\}$. We begin by parametrizing P as follows. Let first $T^* = T_1 \wedge T_2$, and let $\varepsilon = 0$

if $T_1 = T_2$; $\varepsilon = 1$ if $T_1 > T_2$; and $\varepsilon = 2$ if $T_1 < T_2$. Furthermore, let

$$\begin{aligned} P^* & \text{ distribution of } T^*, \\ P^\varepsilon(\cdot | T^*) & \text{ distribution of } \varepsilon \text{ given } T^*, \\ P_1(\cdot | T^*, 1) & \text{ distribution of } T_1 \text{ given } T^*, \varepsilon = 1, \\ P_2(\cdot | T^*, 1) & \text{ distribution of } T_2 \text{ given } T^*, \varepsilon = 2. \end{aligned}$$

To incorporate the censoring model, let $\tilde{C} = (C_1, C_2)$ be a pair of censoring variables, independent of $\tilde{T}$ and with distribution G . We assume that for every (x, y) in $\mathbb{R} \times \mathbb{R}$ there is an (x', y') with positive G -probability such that $x' > x$, $y' > y$.

As before we only observe

$$(\tilde{Z}, \tilde{\Delta}) = (Z_1, \Delta_1, Z_2, \Delta_2),$$

where $Z_j = T_j \wedge C_j$ and $\Delta_j = I\{T_j \leq C_j\}$. We parametrize the distribution P^Z of $(\tilde{Z}, \tilde{\Delta})$ as follows:

$$\begin{aligned} Z^* &= Z_1 \wedge Z_2 = (T_1 \wedge T_2) \wedge (C_1 \wedge C_2), \\ \Delta^* &= I\{T^* \leq C^*\} \text{ where } C^* = C_1 \wedge C_2, \\ \eta &= 0 \quad \text{if } Z_1 = Z_2, \\ \eta &= 1 \quad \text{if } Z_1 > Z_2, \\ \eta &= 2 \quad \text{if } Z_1 < Z_2. \end{aligned}$$

At the outset it is not clear that Δ^* is observable. The following lemma shows that Δ^* is indeed a function of $(\tilde{Z}, \tilde{\Delta})$.

Lemma 3.1. *The following relation holds:*

$$\{\Delta^* = 1\} = \{\eta = 1, \Delta_2 = 1\} \cup \{\eta = 2, \Delta_1 = 1\} \cup \{\eta = 0, \Delta_1 \vee \Delta_2 = 1\}.$$

Proof. Suppose $\Delta^* = 1$, i.e. $T_1 \wedge T_2 \leq C_1 \wedge C_2$. If $\eta = 1$ then $T_1 \wedge C_1 > T_2 \wedge C_2$. We need to show that $\Delta_2 = 1$. Suppose the contrary, i.e. $T_2 > C_2$; then $T_2 \wedge C_2 = C_2$ and we have

$$T_1 \wedge C_1 > C_2 \rightarrow T_1 > C_2$$

which gives $T_1 \wedge T_2 > C_2$ and hence $T_1 \wedge T_2 > C_1 \wedge C_2$, contradicting $\Delta^* = 1$. A similar argument with $\eta = 2$ shows that the left-hand side is contained in the right-hand side.

Next, suppose $\eta = 0$, $\Delta_1 = 1$, i.e. $T_1 \wedge C_1 = T_2 \wedge C_2$ and $T_1 \leq C_1$. These two easily yield $T_1 \leq C_1 \wedge C_2$. Now, $\Delta_1 = 1$ implies $T_1 \wedge C_1 = T_1$ and together with $\eta = 0$ gives $T_1 \wedge T_2 = T_1$. Hence, $\Delta^* = 1$. Similarly if $\eta = 1$, $\Delta_2 = 1$, i.e. $T_1 \wedge C_1 > T_2 \wedge C_2$ and $T_2 \wedge C_2 = T_2$. Then $T_1 \wedge C_1 > T_2$ so that $T_2 < C_1$. All this results in

$$T_2 \leq C_1 \wedge C_2 \text{ implying } T_1 \wedge T_2 \leq C_1 \wedge C_2 \text{ implying } \Delta^* = 1.$$

Similar arguments handle the other cases. $\square$

Turning to the distribution of $(\tilde{Z}, \tilde{\Delta})$, $P^{\tilde{Z}, \tilde{\Delta}}$ can be written as a product

$$P^{(Z^*, \Delta^*)} P^\eta(\cdot | (Z^*, \Delta^*)) P^{Z_\eta}(\cdot | (Z^*, \Delta^*), \eta),$$

where $Z_0 = Z^*$. We will show next that P can be recovered from

$$P^{(Z^*, \Delta^*)}, \quad P^\eta(\cdot | (Z^*, \Delta^* = 1)), \quad P^{Z_\eta}(\cdot | (Z^*, \Delta^* = 1), \eta).$$

(1) The relation between T^* and (Z^*, Δ^*) is obvious. This is just a one-dimensional censoring model where T^* is censored by the independent variable C^* . This gives

$$P\{T^* = t | T^* \geq t\} = \frac{P\{Z^* = t, \Delta^* = 1\}}{P\{Z^* \geq t\}}.$$

(2) For t with $P\{T^* = t\} > 0$,

$$P\{\varepsilon = 1 | T^* = t\} = P\{\eta = 1 | Z^* = t, \Delta^* = 1\},$$

$$P\{\varepsilon = 2 | T^* = t\} = P\{\eta = 2 | Z^* = t, \Delta^* = 1\},$$

$$P\{\varepsilon = 0 | T^* = t\} = P\{\eta = 0 | Z^* = t, \Delta^* = 1\}.$$

This follows from

$$\begin{aligned} P\{\varepsilon = 1 | T^* = t\} &= \frac{P\{\varepsilon = 1, T^* = t\}}{P\{T^* = t\}} \\ &= \frac{P\{\varepsilon = 1, T_1 > t, T_2 = t, C_1 \geq t, C_2 \geq t\}}{P\{T^* = t, C^* \geq t\}} \\ &= \frac{P\{\eta = 1, Z^* = t, \Delta^* = 1\}}{P\{Z^* = t, \Delta^* = 1\}}, \end{aligned}$$

the other cases are handled similarly.

(3) For any E , let P_E stand for the conditional probability given E . For $t > j$,

$$P_{T^*=j, \varepsilon=1}\{T_1 = t | T_1 \geq t\} = \frac{P_{Z^*=j, \Delta^*=1, \eta=1}\{Z_1 = t, \Delta_1 = 1\}}{P_{Z^*=j, \Delta^*=1, \eta=1}\{Z_1 \geq t\}}.$$

To put it differently, what we have is the following one-dimensional censoring model. Consider random variables $\hat{T}$ distributed as $P_{T^*=j, \varepsilon=1}$, and $\hat{C}$ independent of $\hat{T}$ distributed as $P^{C_2}\{\cdot | C_2 > j\}$. Then the pair $\hat{Z} = \hat{T} \wedge \hat{C}$, $\hat{\Delta} = I\{\hat{T} \leq \hat{C}\}$ has the same distribution as (Z_1, Δ_1) under $P_{Z^*=j, \Delta^*=1}$. The details are as follows:

$$\begin{aligned} P_{T^*=j, \varepsilon=1}\{T_1 = t | T_1 \geq t\} &= \frac{P\{T_1 = t, T^* = j, \varepsilon = 1\}}{P\{T_1 \geq t, T^* = j, \varepsilon = 1\}} \\ &= \frac{P\{T_1 = t, T^* = j, C_1 \geq t, C_2 \geq j\}}{P\{T_1 \geq t, T^* = j, C_1 \geq t, C_2 \geq j\}} \\ &= \frac{P\{Z_1 = t, \Delta_1 = 1, Z^* = j, \Delta^* = 1, \eta = 1\}}{P\{Z_1 \geq t, \Delta_1 = 1, Z^* = j, \Delta^* = 1, \eta = 1\}} \\ &= \frac{P_{Z^*=j, \Delta^*=1, \eta=1}\{Z_1 = t, \Delta_1 = 1\}}{P_{Z^*=j, \Delta^*=1, \eta=1}\{Z_1 \geq t, \Delta_1 = 1\}}. \end{aligned}$$

(4) Similarly,

$$P_{T^*=j, \varepsilon=2}\{T_2 = t | T_1 \geq t\} = \frac{P_{Z^*=j, \Delta^*=1, \eta=2}\{Z_2 = t, \Delta_2 = 1\}}{P_{Z^*=j, \Delta^*=1, \eta=2}\{Z_2 \geq t, \Delta_2 = 1\}}.$$

We have thus obtained the distribution of $\tilde{T}$ from that of $(\tilde{Z}, \tilde{\Delta})$. A couple of aspects of this identification deserves mention.

(a) In order to recover the distribution of T^* , as done in the Kaplan–Meier estimate, we use the distribution of (Z^*, Δ^*) .

(b) In obtaining the conditional distributions $P\{T_1 = t \mid T^* = t^*, \varepsilon = 1\}$ we only use factors of the form $P\{\cdot \mid Z^* = t^*, \Delta^* = 1\}$. More specifically terms like $P\{\cdot \mid Z^* = t^*, \Delta^* = 0\}$ are not needed. As mentioned earlier in this section, this amounts to showing that the map $\phi: F \mapsto F(\tilde{Z}, \tilde{\Delta})$ is not onto. This phenomenon reappears when we attempt to write expressions for the likelihood function.

The difficulty in the bivariate case stems from fact that, in general, the empirical distributions are not in the range of ϕ and consequently have no unique inverse. For instance, let $Z_1^*, \dots, Z_k^*$ be observations with $\Delta_i^* = 1, \eta_i = 1, \Delta_{1,i} = 1$. Suppose Z_0 is an observation with $\Delta^* = 0, \eta_0 = 1$ and $\Delta_{0,1} = 0, \Delta_{0,2} = 1$. Assume that $Z_{0,1}, Z_{0,2}$ are distinct from the rest. From the one-dimensional case it follows that P^* will give positive values to $Z_1^*, \dots, Z_k^*$ and $P\{T_1 = Z_{0,1} \mid T_2 = Z_i^*\} = 0$ for $i = 1, \dots, k$. On the other hand, since $\Delta^* = 0$ corresponds to $T_2 > Z_{1,0}^*$, $P\{T_1 = Z_{0,1} \mid T_2 = Z_i^*\} > 0$ for some i .

3.1. The likelihood

The observed random variable has three components, namely (Z^*, Δ^*) , η and (Z_η, Δ_η) . Consequently, the likelihood also has three factors

$$P^{(Z^*, \Delta^*)}, \quad P^\eta(\cdot \mid (Z^*, \Delta^*)), \quad P^{(Z_\eta, \Delta_\eta)}(\cdot \mid (Z^*, \Delta^*), \eta).$$

(1) The distribution of (Z^*, Δ^*) is the same as that coming from the one-dimensional problem, where the survival time is T^* and the censoring variable is C^* . The exact form of $P^{(Z^*, \Delta^*)}$ does not concern us here. We note that $P^{(Z^*, \Delta^*)}$ depends only on P^* .

(2) We have

$$\frac{P\{\eta = j, (Z^* = z, \Delta^* = 1)\}}{P\{Z^* = z, \Delta^* = 1\}} = \frac{P\{\eta = j, T^* = z, C^* \geq z\}}{P\{Z^* = z, C^* \geq z\}} = \frac{P\{\varepsilon = j, T^* = z\}}{P\{T^* = z\}}.$$

Hence,

$$P\{\eta = j \mid Z^* = z, \Delta^* = 1\} = P\{\varepsilon = j \mid t^* = z\}.$$

(3) We next move to the case $P^{(Z_\eta, \Delta_\eta)}(\cdot \mid Z^*, \Delta^* = 1, \eta)$. This again corresponds to a one-dimensional model. Let $\hat{T}$ have the distribution $P^{T_1}\{\cdot \mid T^* = z^*, \varepsilon = 1\}$ and $\hat{C}$ be independent of $\hat{T}$ have the distribution $P^{C_2}\{\cdot \mid C_1 > z^*, C_2 \geq z^*\}$. Then given $Z^* = z^*, \Delta^* = 1, \eta = 1, (Z_1, \Delta_1)$ has the same distribution as $\hat{T} \wedge \hat{C}, I\{\hat{T} \leq \hat{C}\}$. More explicitly,

$$\begin{aligned} \frac{P\{Z_1 = z, \Delta_1 = \delta_1, Z^* = z^*, \Delta^* = 1, \eta = 1\}}{P\{Z^* = z^*\}} &= \frac{P\{T_1 = z, C_1 \geq i, T_2 = z^*, C_2 \geq z^*\}}{P\{T_1 > z^*, T_2 = z^*, C_1 > z^*, C_2 \geq z^*\}} \\ &= \frac{P\{T_1 = z, T_2 = z^*\}}{P\{T_1 > z^*, T_2 = z^*\}} \frac{P\{C_1 \geq z, C_2 \geq z^*\}}{P\{C_1 > z^*, C_2 \geq z^*\}}. \end{aligned}$$

The $\Delta_1 = 0$ cases are handled similarly.

What is left are the terms with $\Delta^* = 0$, namely

$$P^\eta\{\cdot \mid \Delta^* = 0\}, \quad P^{(Z_\eta, \Delta_\eta)}\{\cdot \mid \Delta^* = 0, \eta\}.$$

The first of these depends only on G and the second does not have an explicit expression in terms of the parametrization of P that we have used.

4. Prior and posterior

We first look at the time-discrete case, where P and G are distribution with supports in $\{1, 2, 3, \dots\} \times \{1, 2, 3, \dots\}$. A prior Π for P is called a Beta process, see Hjort (1990), if:

- (a) The hazards $P\{T^* = i \mid T^* \geq i\}$ are independent and are distributed as Beta random variables with parameters $(\alpha^*(i), \beta^*(i))$.
- (b) $P^\varepsilon\{\cdot \mid T^* = t\}$ has a Dirichlet distribution with parameters $(\alpha(0 \mid t), \alpha(1 \mid t), \alpha(2 \mid t))$.
- (c) For $\varepsilon = 1, 2$, the hazards

$$P\{T_\varepsilon = i + d \mid T^* = i, \varepsilon\} / P\{T_\varepsilon \geq i + d \mid T^* = i, \varepsilon\}$$

are independent Betas with parameters $(\alpha_{i,\varepsilon}(d), \beta_{i,\varepsilon}(d))$.

- (d) The families appearing in (a)–(c) are independent.

With i.i.d. observations $(\tilde{Z}_1, \tilde{\Delta}_1), \dots, (\tilde{Z}_n, \tilde{\Delta}_n)$, the likelihood factors as follows:

- (a) Probabilities of $(Z_1^*, \Delta_1^*), \dots, (Z_n^*, \Delta_n^*)$. These arise from a one-dimensional censoring model involving P^* and G .
- (b) Probabilities of η from Z_i^* s with $\Delta_i^* = 1$.
- (c) Probabilities of Z_j from Z_i^* s with $\Delta_i^* = 1$ and $\eta = j$ for $j = 1, 2$.
- (d) Probabilities arising from Z_i^* s with $\Delta_i^* = 0$. These do not have an explicit expression in terms of the parameters $P^*, P^\varepsilon\{\cdot \mid T^*\}$, etc.

In order to arrive at a tractable ‘posterior’ we will ignore terms in ‘(d)’ above and work with an *incomplete likelihood* involving only (a)–(c). We view the terms related to (a)–(c) as more statistically relevant, for the estimation of the bivariate survival curve, than those related to (d). If this is done then it is easy to see that the resulting ‘posterior’ is again a Beta process. Somewhat more formally, assume for simplicity that $\eta_i \neq 0$ for $i = 1, \dots, n$, and set

$$\begin{aligned} Y_i^* &= \#\{j: Z_j^* \geq i\}, \\ \Delta N_i^* &= \#\{j: Z_j^* = i, \Delta_j^* = 1\}, \\ N_i(1) &= \#\{j: Z_j^* = i, \Delta_j^* = 1, \eta = 1\}, \\ N_i(2) &= \#\{j: Z_j^* = i, \Delta_j^* = 1, \eta = 2\}, \\ Y_{i,d}(1) &= \#\{j: Z_j^* = i, \Delta_j^* = 1, \eta = 1, Z_{1,j} \geq i + d\}, \\ \Delta N_{i,d}(1) &= \#\{j: Z_j^* = i, \Delta_j^* = 1, \eta = 1, Z_{1,j} = i + d, \Delta_{1,j} = 1\}, \\ Y_{i,d}(2) &= \#\{j: Z_j^* = i, \Delta_j^* = 1, \eta = 2, Z_{2,j} \geq i + d\}, \\ \Delta N_{i,d}(2) &= \#\{j: Z_j^* = i, \Delta_j^* = 1, \eta = 2, Z_{2,j} = i + d, \Delta_{2,j} = 1\}. \end{aligned}$$

The updated parameters are

$$\begin{aligned}\hat{\alpha}^*(i) &= \alpha^*(i) + \Delta N_i^*, \\ \hat{\beta}^*(i) &= \beta^*(i) + Y_i^* - \Delta N_i^*, \\ (\hat{\alpha}(0|i), \hat{\alpha}(1|i), \hat{\alpha}(2|i)) &= (\alpha(0|t) + N_i(0), \alpha(1|t) + N_i(1), \alpha(2|t) + N_i(2)), \\ \hat{\alpha}_{i,1}(d) &= \alpha_{i,1}(d) + \Delta N_{i,d}(1), \\ \hat{\beta}_{i,1}(d) &= \beta_{i,1}(d) + Y_{i,d}(1) - \Delta N_{i,d}(1), \\ \hat{\alpha}_{i,2}(d) &= \alpha_{i,2}(d) + \Delta N_{i,d}(2), \\ \hat{\beta}_{i,2}(d) &= \beta_{i,2}(d) + Y_{i,d}(2) - \Delta N_{i,d}(2).\end{aligned}$$

An easy computation gives the Bayes estimates as

$$\begin{aligned}E_n P\{T^* = i | T^* \geq i\} &= \frac{\alpha^*(i) + \Delta N_i^*}{\alpha^*(i) + \beta^*(i) + Y_i^*}, \\ E_n P\{T^* > i | T^* \geq i\} &= 1 - \frac{\alpha^*(i) + \Delta N_i^*}{\alpha^*(i) + \beta^*(i) + Y_i^*} \\ &= \frac{\beta^*(i) + Y_i^* - \Delta N_i^*}{\alpha^*(i) + \beta^*(i) + Y_i^*}, \\ E_n P^e\{j | T^* = i\} &= \frac{\alpha(j|t) + N_i(j)}{\alpha(0|t) + N_i(0) + \alpha(1|t) + N_i(1) + \alpha(2|t) + N_i(2)}, \\ E_n \frac{P\{T_1 = i + d | T^* = i, \varepsilon = 1\}}{P\{T_1 \geq i + d | T^* = i, \varepsilon = 1\}} &= \frac{\alpha_{i,1}(d) + \Delta N_{i,d}(1)}{\alpha_{i,1}(d) + \beta_{i,1}(d) + Y_{i,d}(1)}, \\ E_n \frac{P\{T_1 > i + d | T^* = i, \varepsilon = 1\}}{P\{T_1 \geq i + d | T^* = i, \varepsilon = 1\}} &= 1 - \frac{\alpha_{i,1}(d) + \Delta N_{i,d}(1)}{\alpha_{i,1}(d) + \beta_{i,1}(d) + Y_{i,d}(1)} \\ &= \frac{\beta_{i,1}(d) + Y_{i,d}(1) - \Delta N_{i,d}(1)}{\alpha_{i,1}(d) + \beta_{i,1}(d) + Y_{i,d}(1)}.\end{aligned}$$

We also have

$$\begin{aligned}P\{T_1 = i + d, T_2 = i\} \\ &= \left[\prod_1^{i-1} \frac{P\{T^* > j\}}{P\{T^* \geq j\}} \right] \frac{P\{T^* = i\}}{P\{T^* \geq i\}} P\{\varepsilon = 1 | T^* = i\} \\ &\quad \times \left[\prod_1^{d-1} \frac{P\{T_1 > i + d' | T^* = i, \varepsilon = 1\}}{P\{T_1 \geq i + d' | T^* = i, \varepsilon = 1\}} \right] \left[\frac{P\{T_1 = i + d | T^* = i, \varepsilon = 1\}}{P\{T_1 \geq i + d | T^* = i, \varepsilon = 1\}} \right]. \quad (4.1)\end{aligned}$$

As we let the prior parameters go to zero, for each j , the Bayes estimate of $P\{T^* > j\}/P\{T^* \geq j\}$ converges to N_j^{*+}/N_j^* , where $N_j^{*+} = Y_j^* - \Delta N_j^*$ and the Bayes estimate of $P\{T^* = i\}/P\{T^* \geq i\}$ converges to $\Delta N_i^*/Y_i^*$. Carrying out a similar evaluation of the limit in each of the factors in (4.1) and setting $N_{i,d'}^+(1) = Y_{i,d'}(1) - \Delta N_{i,d'}(1)$ we get the ‘noninformative’ Bayes estimator

$$\left\{ \prod_1^{i-1} \frac{N_j^{*+}}{N_j^*} \right\} \frac{\Delta N_i^*}{Y_i^*} \frac{N_i(1)}{\Delta N_i^*} \left\{ \prod_1^{d-1} \frac{N_{i,d'}^+(1)}{N_{i,d'}(1)} \right\} \frac{\Delta N_{i,d}(1)}{Y_{i,d}(1)}.$$

While we have not worked out the details, we do expect this noninformative Bayes estimator to be consistent.

5. Beta processes

In the context of general Beta processes it is convenient to work with the cumulative hazard function of a distribution rather than the distribution function itself. These processes were introduced in Hjort (1990). Recall that if F is a distribution function and if $\bar{F} = 1 - F$ is the survival function then the cumulative hazard function A is the function

$$A(t) = \int_{[0,t]} \frac{dF(s)}{F[s, \infty)}.$$

This is right continuous and increasing, has the same jump points as F , and if t is a jump point then the jump size is $\Delta A(t) = \Delta F(t)/F[t, \infty)$.

Let F_0 be a distribution function and let A_0^d, A_0^c be the discrete and continuous parts of its cumulative hazard function A_0 . Let w be a continuous positive function on $\mathbb{R}^+$. A Beta process with parameters w and A_0 is a process such that the process $\{A(t): t > 0\}$ has independent increments and has the Lévy measure

$$dL(t, s) = w(t)s^{-1}(1-s)^{w(t)-1} dA_0^c(t) ds \quad \text{for } t \in \mathbb{R}^+, \quad 0 \leq s \leq 1$$

and if t_0 is a jump point then

$$\Delta A(t_0) \sim \text{Beta}\{w(t_0)A_0^d(t_0), w(t_0)(1 - A_0^d(t_0))\}.$$

If we have n i.i.d. observations with some of them right censored and if we set

$$\Delta N(t) = \#\{i: Z_i = t, \Delta_i = 1\}, \quad Y(t) = \#\{i: Z_i \geq t\},$$

then Hjort (1990) has shown that the posterior is again a Beta process with parameters $w + Y$ and $\int_0^\cdot w dA_0 + \Delta N)/(w + Y)$.

Since our approach to the bivariate problem is a sequence of one-dimensional censoring models, a natural generalization of the one-dimensional Beta process suggests itself. Let F_0 be a distribution function on $\mathbb{R}^+ \times \mathbb{R}^+$. Let F_0^* be the distribution of T^* , $F_0^{\varepsilon}(\cdot | T^*)$ be the distribution of ε given T^* and $F_0^1(\cdot | T^*)$ the distribution of T_1 given $(T^*, \varepsilon = 1)$ and $F_0^2(\cdot | T^*)$ the distribution of T_2 given $\{T^*, \varepsilon = 2\}$. Next, let w be a function on $\mathbb{R}^+ \times \mathbb{R}^+$. We write $w^*(t) = w(t, t)$. Furthermore if $s > t$, then set $w_1(s | t) = w(t, s)$. Define $w_2(s | t)$ similarly.

Let α_0 be a function from $\mathbb{R}$ to $(\mathbb{R})^3$. The parameter space is

- (1) $\mathcal{M}$ —the set of all distributions on $\mathbb{R}^+$. We view this as the set of all possible distributions for T^* .
- (2) Γ —all measurable functions from $\mathbb{R}$ to the set of probability measures on $\{0, 1, 2\}$. This correspond to the set of conditional distributions of ε given T^* .
- (3) Γ_1 —all measurable functions from $\mathbb{R}$ to $\mathcal{M}$, viewed as the set of distributions for T_1 given T^* and $\varepsilon = 1$.
- (4) Γ_2 —all measurable functions from $\mathbb{R}$ to $\mathcal{M}$, viewed as the set of distributions for T_2 given T^* and $\varepsilon = 2$.

A prior $\Pi = \Pi_1 \times \Pi_2 \times \Pi_3 \times \Pi_4$ on $\mathcal{M} \times \Gamma \times \Gamma_1 \times \Gamma_2$ is called a bivariate Beta process prior if

- (1) Π_1 is a Beta process with parameters C^* , A_0^* on $\mathcal{M}$.
- (2) Under Π_2 , $\{\gamma(t^*, \cdot): t^* \in \mathbb{R}\}$ are independent and for each t^* , $\gamma(t^*, \cdot)$ has a Dirichlet distribution with parameters $\alpha_0(t^*)$, $\alpha_1(t^*)$, $\alpha_2(t^*)$.
- (3) Under Π_3 , $\{\gamma_1(t^*, \cdot): t^* \in \mathbb{R}\}$ are independent and for each t^* , $\gamma_1(t^*, \cdot)$ is a one-dimensional Beta process with parameters $w_1(\cdot | t^*)$ and $A_0^{(1)}(\cdot | t^*)$.
- (4) Under Π_4 , $\{\gamma_2(t^*, \cdot): t^* \in \mathbb{R}\}$ are independent and for each t^* , $\gamma_2(t^*, \cdot)$ is a one-dimensional Beta process with parameters $w_2(\cdot | t^*)$ and $A_0^{(2)}(\cdot | t^*)$.

A technical argument is needed to properly demonstrate the existence of the above process; see the Appendix.

It is fairly routine to see that given n i.i.d. observations, using the incomplete likelihood, the posterior is again a bivariate Beta process. The independence assumption in the prior and the product form of the likelihood ensures the required independence for the posterior. That the posterior is again a Beta process follows from the one-dimensional results. The exact Bayes estimators can be written down but become notationally tedious. We just note that if we let w go to zero we get exactly the same noninformative-prior estimator as before.

6. An example

We first examine [Pruitt's \(1988\)](#) example via our approach. The only observations with $\Delta^* = 1$ are the ones with both coordinates uncensored. Thus, the estimate of the distribution of T^* will have positive mass only at Z^* corresponding to these points. The censored observations will all have $Z^* = 0$. Also it is only the observations with both coordinates uncensored that are used to estimate the conditional probabilities.

Considering a simple numerical example in [Dabrowska \(1988\)](#), suppose that the observed points written as $(Z_1, Z_2, \Delta_1, \Delta_2)$ are

$$(.51, .02, 1, 1), (.11, .62, 1, 0), (.24, .24, 0, 0), (.68, .68, 1, 1).$$

The noninformative-prior estimate will give mass $\frac{1}{4}$ to each of $(.11, .62)$, $(.51, .02)$ and $\frac{1}{2}$ to the point $(.68, .68)$.

Estimate of	Dabrowska	Noninformative-prior
$P\{T_1 > 0, T_2 > 0\}$	1	1
$P\{T_1 > .11, T_2 > 0\}$	.75	.75
$P\{T_1 > .24, T_2 > 0\}$	.75	.75
$P\{T_1 > .51, T_2 > 0\}$	.375	.5
$P\{T_1 > .68, T_2 > 0\}$	0	0
$P\{T_1 > 0, T_2 > .02\}$	.75	.75
$P\{T_1 > 0, T_2 > .24\}$	.75	.75
$P\{T_1 > 0, T_2 > .62\}$	.75	.5
$P\{T_1 > 0, T_2 > .68\}$	0	0
$P\{T_1 > x, T_2 > y\}$	.5	.5

In the last row $x \in [.11, .68]$ and $y \in [.02, .68]$.

The set in the first column of the fourth row contains that in the last row, for $x > .51$. However, using the Dabrowska estimator method, the set in the last row has larger probability, leading to negative mass for some sets. Pruitt (1991) has examined the occurrence of negative mass in the Dabrowska estimate. This lack of monotonicity will not occur with the noninformative estimator.

Appendix A.

Consider Γ_1 . This is a subset of $\mathcal{M}^{\mathbb{R}}$. Kolmogorov's consistency theorem guarantees the existence of a product measure on $\mathcal{M}^{\mathbb{R}}$. However, we need to show that the outer measure of Γ_1 is 1. This follows from the following lemma. A similar argument appears in Judd (1985).

Lemma A.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be a complete separable metric spaces with their corresponding Borel σ -fields. Let $\mu = \prod_{i \in \mathcal{Y}} \mu_i$ be a product of probability measures on $\mathcal{X}^{\mathcal{Y}}$. If $\Gamma \subset \mathcal{X}^{\mathcal{Y}}$ is the set of all measurable functions from $\mathcal{Y}$ to $\mathcal{X}$, then the μ outer measure of Γ is one.*

Proof. Let E be any measurable set in $M(\mathcal{Y})^{\mathcal{X}}$ containing Γ . We will show that E is the whole space. Since E is measurable there is a countable set $\{t_1, t_2, \dots\}$ such that E is in the σ -algebra generated by the projections to these coordinates. Hence, for any γ_1, γ_2 , if $\gamma_1(\cdot | t_i) = \gamma_2(\cdot | t_i)$ for $i = 1, 2, \dots$, then either both γ_1 and γ_2 or none of them belong to E . Take any $\gamma \in M(\mathcal{Y})^{\mathcal{X}}$, and pick a γ_0 from Δ . Define γ^* by

$$\gamma^*(\cdot | t_i) = \begin{cases} \gamma(\cdot | t_i) & \text{for } i = 1, 2, \dots, \\ \gamma_0(\cdot | t) & \text{otherwise.} \end{cases}$$

Since γ_0 is modified only at countably many points, measurability still holds and hence $\gamma^* \in \Gamma$. Now, $\gamma^*(\cdot | t_i) = \gamma(\cdot | t_i)$ for $i = 1, 2, \dots$, hence γ is also in E . $\square$

References

- Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N., 1993. Statistical Models Based on Counting Processes. Springer, New York.
- Dabrowska, D.M., 1988. Kaplan–Meier estimate on the plane. Ann. Statist. 15, 1475–1489.
- Hjort, N.L., 1990. Nonparametric Bayes estimators based on Beta processes in models for life history data. Ann. Statist. 18, 1259–1294.
- Judd, K.L., 1985. The law of large numbers with a continuum of IID random variables. J. Econom. Theory 35, 19–25.
- Langberg, N.A., Shaked, M., 1982. On the identifiability of multivariate life distribution functions. Ann. Probab. 10, 773–779.
- Peterson, A.V., 1977. Expressing the Kaplan–Meier estimator as a function of empirical subsurvival functions. J. Amer. Statist. Assoc. 72, 854–858.
- Pruitt, R.C., 1988. An inconsistent Bayes estimate in bivariate survival curve analysis. University of Minnesota, Preprint 529.
- Pruitt, R.C., 1991. On negative mass assigned in the bivariate Kaplan–Meier estimator. Ann. Statist. 19, 443–453.
- Pruitt, R.C., 1993. Identifiability of bivariate survival curves from censored data. J. Amer. Statist. Assoc. 88, 573–579.